

## Aberystwyth University

### *Archives context and discovery*

Higgins, Sarah; Hilton, Christopher; Dafis, Lyn

*Publication date:*  
2014

*Citation for published version (APA):*

Higgins, S., Hilton, C., & Dafis, L. (2014). *Archives context and discovery: Rethinking arrangement and description for the digital age*. Paper presented at 2nd Annual Conference of the International Council on Archives, Girona, Spain.

#### **Document License** CC BY-NC-SA

#### **General rights**

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)

## Archives context and discovery: rethinking arrangement and description for the digital age

Sarah Higgins

Department of Information Studies, Aberystwyth University

Christopher Hilton

Wellcome Library

Lyn Dafis

National Library of Wales

### Abstract

Traditionally the archival principles of provenance, *respect des fonds* and original order are enacted through hierarchical arrangement and description, facilitating intellectual and physical access and the preservation of context. Access to archival collections is now impeded by the pragmatic impossibility of achieving the ideal of fully arranged collections described in detail to individual item level. Meanwhile online description to fonds or series level - created using standards created for the paper paradigm - ignore the new reality of facilitating access to born-digital material and digitised collections. Both require item level description, the former already incorporating its own pre-packaged metadata, which enables the reader to access and use digital material without necessarily knowing their context and provenance. Both, too, can be arranged in multiple ways. New paradigms for arrangement and description for the digital age need to focus on individual items and the user experience. Opportunities to facilitate contextual understanding and access include: user generated arrangement and description, tagging and linkage to existing biographical, historical and contextual resources. This study, a collaboration between the UK organisations: Aberystwyth University, the National Library of Wales and the Wellcome Library, proposes a user study that will offer archival collections digitally with no pre-defined archival arrangement and minimal contextual information. Tools will be provided to enable user input to both the arrangement and description of the collections and behaviours will be analysed to help identify and evaluate new methods for enabling access and maintaining contextual information.

### Archival principles

In Europe and the United States, archival processes of appraisal (selection), arrangement and description have been underpinned by the principles of provenance, *respect des fonds* and original order since the publication of the *Dutch manual* in the late 1800s (Barrit, 1993; Muller, Feith, & Fruin, 1968). Adhering to these principles when processing an archival collection enables the archivist to maintain the context of the creation and original use of the materials while ensuring that the essential characteristics of records (and by extrapolation archival materials) are maintained: authenticity, reliability, usability and integrity (International Organization for Standardization, 2001, p.7). Arrangement and description of collections supports both their physical and intellectual management - Jenkinson's *physical and moral defence of archives* (Jenkinson, 1966, pp. 44-105). These processes embody the skill of the archivist - interpreting and representing collections through analytical processes to arrive at "an understanding of the records sufficient for contextualising and providing intellectual access to them" (Meehan, 2009, p. 73). Hierarchical descriptive practices – where the *fonds*

is the primary unit of control - have developed in response to archival principles, so that “the records being described should represent a distinct and coherent whole, one that will illuminate, and not obscure, the context of activities out of which the records were created and maintained during their active life” (MacNeil, 2002, p.202). As a profession, archivists have been confident in their ability to record “objective truth” regarding the context of the creation and use of records selected for retention: deciding what information regarding this will be included in the descriptions, and the level of detail required. However, archival principles are rooted in a particular time when Jenkinson’s *defence of the archives* (1966) was an inward looking process – concentrating on safeguarding archival materials rather than facilitating access to them. Description and arrangement now needs to support two functions: recording context through implementing archival principles and enabling the discovery of archival materials through accessible cataloguing.

The tension between these two functions, and how to rationalise them within the confines of archival principles, is not a new issue for archivists. It was being debated, particularly in the realm of digital materials, 20 years ago (Duff, 1995; MacNeil, 1994, 1995; Wallace, 1995). More recently Bunn (2013) urged the archives profession to consider anew how current practice can distinguish between the inward looking focus of archival control and the outward looking focus, required by modern archives, of information discovery and data exchange. Yeo (2012) urges the profession to reconsider original order and “take advantage of some of the new capabilities [of new technology] to help overcome some of the contextual limitations of hierarchical classification schemes and paper-world methods” (p.61) to enable the user to create bespoke arrangements of digital archival materials, that reflect their own interests or world-view.

The short-comings of the lengthy process of archival arrangement practice, in enabling access for the large collections created in the digital age (but not considering digital collections *per se*), was highlighted in Greene and Meissner’s influential article *More product, less process: pragmatically revamping traditional processing approaches to deal with late 20<sup>th</sup>-century collections* (2005). They noted that despite processing backlogs “the archival profession has been unwilling or unable to change its processing benchmarks in response to the greater quantities of acquisitions” (p.1) essentially denying access to hundreds of collections. They called for the implementation of tough appraisal decisions and a move away from item level processing and cataloguing, along with the production of detailed hierarchical finding-aids, reminding us that “researchers are coming to do research, so you don’t have to do it for them in advance.... Use your time wisely” (p.8). McCrea (2006) testifies how top-down processing enables collections to be “dealt with” more quickly, while description at only higher levels of the hierarchical arrangement leaves “more of the burden of discovery on the user rather than on the archive staff” (p. 288).

*More product, less process*, a pragmatic approach for pressured professionals, would seem to speed up the process of arrangement, in the analogue world, while creating finding-aids that, with their lack of detail, reinforce an inward looking focus. More collections are controlled both physically and intellectually, but granular discovery depends on the user researching contextual detail. The logical conclusion of the *more product, less process* approach is that contextual detail will need to be out-sourced to existing accessible resources that can provide the necessary background information, such as Wikipedia, over which the archivist has no control, rather than carefully drafted administrative or biographical histories and authority files.

### **Current standards and access**

The development of personal computers and archival standards such as General International Standard Archival Description (ISAD(G)) (International Council on Archives, 2000) and Rules for Archival Description (RAD) (Planning Committee on Descriptive

Standards, 2008), which codified the structure of archival description, was a trigger point for a more outwardly looking focus to the profession, as there was now a means to share descriptions more widely, both with the user and other organisations. These standards codified archival principles, with *respect des fonds* as “the basis of archival arrangement and description” (Planning Committee on Descriptive Standards, 2008, p. xxiii) with arrangement in an hierarchical model (International Council on Archives, 2000, p. 36). In this they show their North Atlantic origins, taking *respect des fonds* as the key organising principle in the European/North American tradition: they do not so easily accommodate the series model seen in Australian work.

ISAD(G)’s beginning were in the late 1980s with the first draft, and the principles on which it is predicated published in 1992 (International Council on Archives, 1992a, 1992b) and the first edition in 1994 (Ad Hoc Commission on Descriptive Standards, 1994). It emerged at the same time as software started to become available that could render hierarchy meaningfully on-screen and was rapidly adopted without any identifiable widespread pilots. By delineating a small number of standard fields, and identifying a minimum number that must be populated, it made it possible to create metadata that fits easily into a database. Meanwhile its stress upon the key role of the fonds, and the hierarchies that flow from it, enabled the profession *inter alia* to give a clear direction to software suppliers. This ensured that the profession had its own appropriate database software to work with rather than seeking to cannibalise less appropriate library or all-purpose systems. CALM for Archives (a forerunner of Axiell CALM) and Adlib Archive, the two packages that began their domination of the United Kingdom archive sector at this time, boasted ISAD(G) conformance from 1999 (Adlib Information Systems, 1999; DS, 1999). The first edition of Encoded Archival Description (EAD) (Society of American Archivists & Library of Congress, 1998) enabled machine readable mark-up “firmly rooted in archival principles, tradition, and theory” (Library of Congress, 1999) of “a detailed hierarchical analysis of the whole and its sub-components, with an emphasis on provenance and the organization, arrangement and content of material”. Based on ISAD(G) it enabled machine readable semantic interoperability between archival descriptions from different organisations, and was quickly adopted, especially in the UK, where a “small band of early implementers sprang up” (Sweet, 2001) using it as a means of retro-converting paper finding-aids, on an institutional level, while developing hierarchical finding-aids which could be accessed and searched over the Internet (Bouché, 1997; R. Higgins, 1998; Johnston, 2001).

Joint projects and *archival networks* which enabled EAD encoded finding-aids to be shared and delivered from the same portal were equally quick to develop (S. Higgins & Inglis, 2003; Hill et al., 2004). Many of the web-accessible data aggregators created in the UK in the early 2000s, using ISAD(G) to structure data and EAD as their mark-up language, continue to operate, although not all continue to be maintained. These *archival networks* included: large-scale general resources such as the Archives Hub to bring together archives held in higher education establishments (Archives Hub, n.d.-a), or AIM25 (for repositories within London) (AIM25, 2013); thematic, web-based “one-stop shop” sites bringing together information about various repositories’ resources on particular subjects (e.g. Genesis on women’s history – now integrated into the Archives Hub - or Mundus on missionary history) (Archives Hub, n.d.-b; Genesis Project Team, 2004; School of Oriental and African Studies, n.d.); and those bringing together the archives held by particular nations such as The Scottish Archive Network (Scottish Archive Network Ltd, n.d.) and Archives Wales (National Library of Wales, 2014). None of this activity would have been possible without the standards, which unlocked the opportunity presented by the headlong development of the World Wide Web. The dream of a UK *über* network to be managed by The National Archives, that linked all the networks together (including not only finding-aids, but digitised images of archives) (Craven, 2004) was not realised as the necessary funding could not be raised. This was the corner-stone recommendation of *The Archives Taskforce* to “bring about a step change in the accessibility of services and collection” (Archives Task Force, 2004).

Adoption of ISAD(G) and EAD, and of allied standards similarly rooted in archival principles, (for example, RAD in Canada and Describing Archives: A content standard (DACS) in the US) (Society of American Archivists, 2013), therefore enabled innovative creation and sharing of data based on standardised methodologies for creating archive finding-aids. These, in turn, informed software suppliers' work and have transformed not only the job of archival cataloguing but the purpose of the catalogue. The emphasis was shifted from the inward looking *defence* of the archive to the outward looking process of providing information, enabling access and through associated digitisation projects, providing surrogates.

However, fonds-based software has also helped to cement a particular mind-set regarding descriptive practice (Figure 1) so that archival principles and the use of standards derived from these are reinforced. Little has changed in archival descriptive practice, at least in the UK, since the enthusiastic uptake of descriptive standards in the late 1990s.

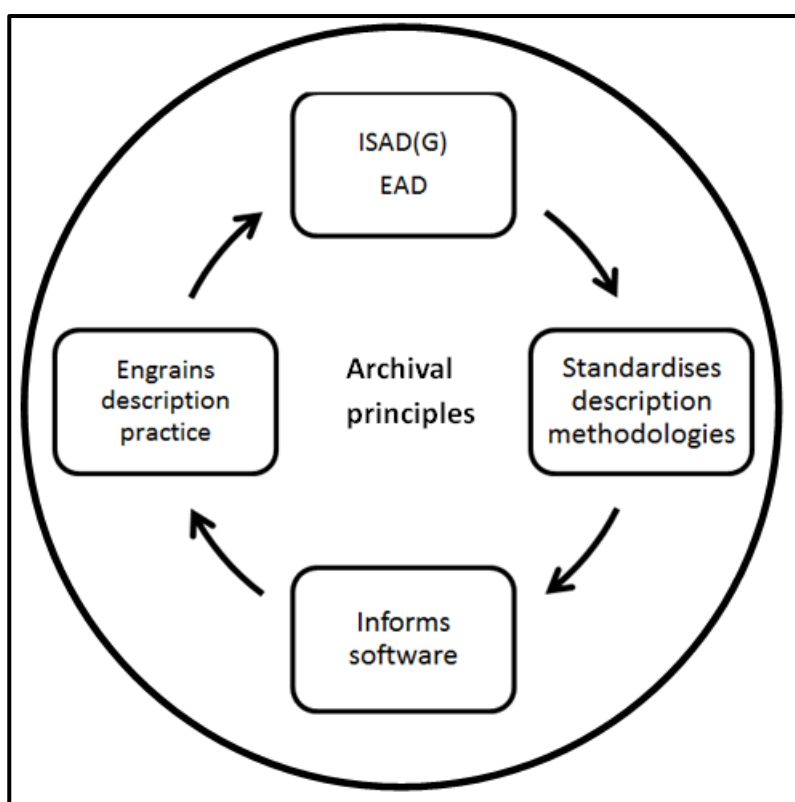


Figure 1: Archival descriptive practice feedback cycle

The paradigm shift that embedded descriptive standards into professional practice, and the impact of this is overwhelmingly positive in terms of the improvements in access brought about by the systematic structuring of both new archival descriptions, the retro-conversion of existing ones; and the subsequent exchange of these descriptions to create online union catalogues. However, they carry with them the danger of groupthink, of rendering it harder, if not impossible to conceptualise alternative ways of seeing, or to explore alternative or supplementary models.

### Access in the digital age

This engraining of descriptive practice has consequences that now impact upon archivists seeking to work in an era of data exchange and interactivity.

- Firstly, enmeshing all catalogue data firmly into a hierarchical catalogue, which does “not repeat information at a lower level of description that has already been given at a higher level” (International Council on Archives, 2000, p. 21) reduces the extent to which a description of an individual series, file or item can make sense divorced from the context of the fonds.
- Secondly, it privileges one particular set of relationships for each individual fonds, series, file or item description whilst downplaying others.

It is easy to forget that ISAD(G), and by association EAD, represent the work of a particular historical moment. Looking back with hindsight, ISAD(G) is clearly rooted in the 1990s and the beginning of the information revolution (although a perfectly ISAD(G) compliant catalogue can be produced using paper and pencil). Its assumptions are very much of that Web 1.0 era: where information was presented in a way which mimicked paper; and was served up on the Web for people to passively consume. Meanwhile, because of its solid roots in archival principles (and the professional archivist’s outlook), its validity has not been seriously challenged.

But more recent developments have changed the way people engage and more importantly interact with the Internet. These include: the improvement in Web browser search algorithms, affordable reliable broadband, the advent of mass digitisation, the exponential growth in Web 2.0 and the heralding of Web 3.0. The first two make Internet use and reliable searches an accessible reality. Mass digitisation makes possible immersive interaction with archival material for the user, eliding the distinction in their experience between metadata and primary material: one screen can provide the venue for both the archivist’s description and the raw material. With the addition of Web 2.0 functionality, to enable users to add notes or comments, the Web based experience can bring together the roles previously played by paper catalogue, reading room desk and user’s notebook. Web 3.0 (characterised by the Semantic Web or linked data), currently being enacted in experimental silos, envisages a global database where all information is linked through semantic applications.

Two case studies of archive repositories illustrate where the archival profession is at present and the challenges faced. Both are moving from the production of detailed hierarchical catalogues as the end-product of archival processing, towards a greater emphasis on quick reader-access while linking catalogues to digital facsimiles.

### **Digital access at the National Library of Wales**

The National Library of Wales (NLW) has a broad collecting remit, taking “archives of selected organizations and associations that operate on a national scale in Wales” and “personal papers of writers, artists, musicians, politicians, scholars and others who have played an important part in the life of the nation” (Roberts, 2014). This spreads the net of potential acquisitions very wide and perhaps predictably has resulted in a cataloguing backlog, a scenario familiar to most archive repositories. In summer 2014 an internal strategic look at the backlog and at cataloguing procedures, was undertaken, to see how access might be improved to collections currently condemned to languish uncatalogued for years.

The report of that process (Roberts, 2014) looked at all parts of the Library’s archival workflow, arguing that the backlog might be addressed through more stringent application of the collection development policy: the statement that papers are to be acquired from the sources listed above “having a regard to the content and richness of the archive” had, it was felt, been too often ignored with papers acquired because of the significance of the creator, not the archive itself. In addition to tighter control of what was acquired, a programme of deaccessioning was proposed for material whose research merits were not sufficient for it to

earn its place in the Library's holdings. In between these two poles the report looked at the whole process of cataloguing, the process of arrangement and metadata-creation that was necessary before the public could have access to material. Most notably, the report noted that "since the advent of ISAD(G) and adopting this standard for listing archives, we have not questioned or really evaluated the effectiveness of our cataloguing methods".

Cataloguing practice at NLW has moved away from the very detailed catalogues of earlier years, with description now taking place at the level of the file or equivalent rather than the individual piece of paper; however, it was felt that gains in time from this change had been offset by time-consuming processes such as detailed arrangement, careful drafting of data to ensure compliance with standards (such as those for authority files), justifying exceptions to existing practices, and so forth. The recent report argues for a revision of what was felt to constitute the basic minimum necessary in a catalogue description. It also proposes an approach based in pragmatism, in which the level of detail in a given catalogue is considered on a bespoke, case-by-case basis, rather than the Library aiming for consistency across its collections and adherence to "an ideal or arbitrary standard."

A reduction in the level of detail in catalogue metadata is proposed, and also in the amount of time spent arranging material, in line with the principles of *more product, less process* (Greene & Meissner, 2005) discussed above. The priority henceforward would be to produce a general, collection-level description as soon as possible, "avoiding time-consuming tasks, such as composing a biographical/administration history" (Roberts, 2014), with the aim of making this available through both NLW's own catalogue but also through other resources such as the archives network *Archives Wales* run by the Archives and Records Council Wales (ARCW) (the body that represents archive repositories in Wales and brings together national, local and specialist collections). Crucially, descriptions on *Archives Wales* are held as web-pages accessible to search engines. In order to spread this collection-level description process across as many different collections as possible, lower levels of cataloguing would be made less detailed, and descriptions would be created, if necessary, at the level of the box rather than the file. Such pared down processing runs the risk of failing to identify material for sensitivity and for compliance with Data Protection legislation – meaning that closer liaison will be required with donors and depositors to identify this prior to transference to the repository.

Access to archives at NLW, then, is seen to require the production of metadata, rather than description *per se*, but on a scale less detailed (and possibly less hierarchical) than before. The user is not seen as needing a highly granular catalogue in order to gain access to collections: rather, accessibility demands only that there should be *some* form of catalogue, which may be minimal but should be capable of widespread dissemination. Lightweight catalogues spread widely across NLW's collections and disseminated to the maximum number of users, including remote ones, are seen as offering better service to the reader than the carefully-constructed, detailed catalogues for a small number of collections that would result from previous methods. The production of large numbers of these more lightweight catalogues is seen as a necessary part of the development of a wider audience, wider in the senses both of user communities beyond those currently using the Library and also of geographic distribution. The latter is seen as implying greater demand for digital facsimiles of material to be made available remotely. The use of volunteers and/or crowd-sourcing to enhance catalogues is being considered: the latter, it is implicit, will add detail to individual records, especially for digitised items, but will not affect the arrangement or hierarchical structure of the catalogue.

NLW, then, is moving towards reduction of detail in its catalogues, in terms both of hierarchical arrangement and in amount of description in individual records, reasoning that some access now is better than a collection languishing unseen whilst very detailed catalogues are constructed for the material ahead of it in the cataloguing queue; prioritising

access by search engines and remote readers to these more succinct catalogues; and envisaging that this new approach will be complemented by remote access to digital facsimiles and by volunteer and/or crowd-sourced enhancement of records.

### **Digital access at the Wellcome Library**

The Wellcome Library (WL), one of the world's leading repositories for the history of medicine, has recently sunk considerable effort and resource into a large-scale digitisation programme. This has made a large range of resources accessible to the general public, but has also raised longer-term questions about the usefulness of current archival finding aids. In the first tranche of digitisation, carried out under the heading (Wellcome Library, 2014), the WL digitised all the items in several fonds relating to genetics, including papers of individuals such as the molecular biologist (and co-discoverer of DNA) Francis Crick (1916-2004) as well as organisations involved in questions about heredity such as the controversial Eugenics Society. As part of the same exercise complete allied fonds held at other institutions were also digitised, and the metadata for these ingested into the Wellcome system, enabling seamless searching across various collections. The potential is clear: to take the most obvious example, the papers of Francis Crick at the WL and his collaborator James Watson at Cold Spring Harbor Laboratory are brought virtually together and made available world-wide.

Yet the feedback from users, whilst broadly positive, has demonstrated areas for obvious development in how resources like this might be managed in future. Firstly, digitisation was applied as a bolted-on supplement to what had previously been stand-alone archive catalogues; and secondly, those catalogues had been highly orthodox, hierarchical ones, dating either from the 1980s and 1990s, or from the very early years of the twentieth century when ISAD(G) was first being applied at the WL.

Retro-conversion of the WL's existing finding aids began in 2001 and was completed in 2012. The descriptions fell into two broad categories. Material earlier than 1900 had, in general, been described in a manuscript-cataloguing tradition, with considerable detail in individual descriptions items. These were simply given a running manuscript number, without arrangement in a hierarchy. The content of these detailed descriptions mapped well to ISAD(G) fields, despite predating the standard by fifty years in some cases, and during retro-conversion the opportunity was also taken to bring together and reconstitute hierarchically some collections that had hitherto been disguised by the use only of running manuscript numbers for references. Twentieth century material was catalogued hierarchically, typically with very detailed collection-level descriptions allied to detailed and deep hierarchical structures as a way of organising very succinct descriptions lower down the hierarchy. In both cases, retro-conversion made it possible to search across the entire collection for the first time, without the use of pre-defined access points. Both the use of archive material, and the range of uses to which it was put, increased year on year during the process of retro-conversion.

There is no question that conforming to ISAD(G) enabled conversion of these catalogues to database form and thereby enabled their usefulness to increase greatly. However, several decisions taken at the turn of the century to facilitate quick conversion and word-searchability continue to shape the reader experience now in a very different technical environment:

1. Making the catalogues word-searchable provided quick and good returns for the reader. As a result the further enhancement – linking, disambiguation and so forth – that name authority files would have offered was seen as a low priority.
2. The existing typescript catalogues of twentieth century material relied heavily on hierarchy for readers to understand the context of the material, and expressed this hierarchy through page-layout, rather than explicit verbal content.



Figure 2 shows a portion of the original typescript descriptions for series A of the fonds Medical Women's Federation. As can be seen the file level description (the individual producible unit) only contains three elements: reference number, date and box-number. The context for the creation of the files is held in the title "Council and General Meetings, 1916-1983" of the series level description.

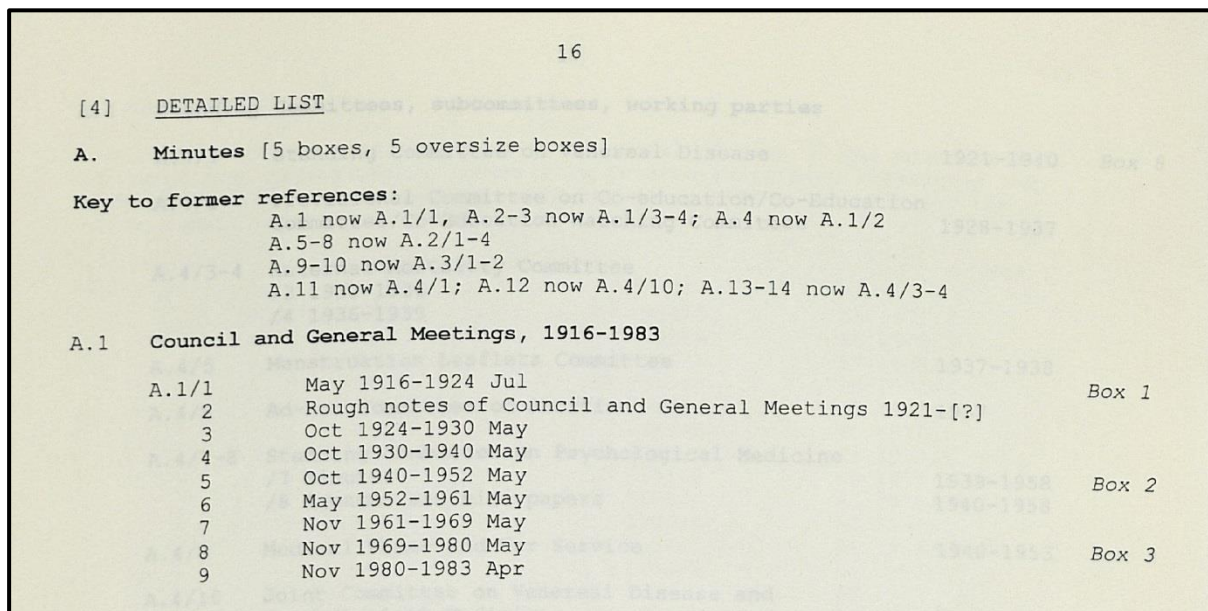


Figure 2: Original typescript finding aid for Series A of the fonds Medical Women's Federation (SA/MWF) held by the Wellcome Library

Retro-conversion to conform to the minimum requirements of the first edition of ISAD(G) (5 units of description: reference code, title, date, level of description and extent) (Ad Hoc Commission on Descriptive Standards, 1994) added some detail to the descriptions at all levels of the hierarchy, but these remain extremely terse (Figure 3).

<b>SAMWF</b>	SA/MWF Medical Women's Federation 1879-1988 Collection
	▼ <b>SA/MWF/A</b> Minutes 1916-1984 Section
	▼ <b>SA/MWF/A.1</b> Council and General Meetings 1916-1983 Series
	<b>SA/MWF/A.1/1</b> Council and General Meetings May 1916-Jul 1924 Item
	<b>SA/MWF/A.1/2</b> Council and General Meetings 1921-[?] Item

Figure 3: Series A of the fonds Medical Women's Federation (SA/MWF) held by the Wellcome Library retro-converted to conform to ISAD(G) version 1.

Following retro-conversion, the individual components of the description became individual database records, currently managed and searched through the Axiell CALM system. As such descriptions can be sorted and ordered individually, and thus can be divorced from the wider context of the fonds. Figure 3 shows the database records ordered to show the collection's hierarchy. The real problems begin when a user's search across the database throws up a series, file or item level description that is divorced from the context of the hierarchy. With such minimalist descriptions, it may not be possible to identify what has been discovered, as there is no information given regarding the overall context. Figure 4 shows how the individual file SA/MWF/A.1/1 (shown in Figure 3) appears the user who surfaces it during a free-text search. There is no way for the user to identify what organisation's "Council and General Meetings" are being referred to, without going into the hierarchy view and navigating up the tree to identify the fonds. It is only by looking at the hierarchy "tree" overall that the reader can gain a clear idea of how this item fits into the fonds as a whole. Feedback from the WL's reading rooms suggests that the need to navigate up or down the hierarchy, to learn an item's context, or to identify the level at which material is producible, is

the single most confusing thing that readers (and non-archivist staff) encounter in the catalogues.

PAGE 1 OF 1

	REFERENCE	LEVEL	TITLE	DATE
1	GC/222/8/26	Piece	<b>The 15th (Scottish) Hospital, Cairo</b>	Mid-late 20th Century
2	MS.6825	Item	<b>Scottish Hospital, London</b>	1799-1843
3	MS.8449	Collection	<b>Lilias Mary Grant and Ethel Moir, nursing orderlies with the Scottish Women's Hospital Serbian/Russian Unit</b>	1916-1972
4	PP/MCS/C/41	Item	<b>CBC representative at Scottish Nursing, Hospital and Health Conference, Edinburgh</b>	1931
5	RAMC/761/3/16	Piece	<b>Report re occupational therapy for patients in No. 15 (Scottish) General Hospital</b>	May 1942
6	RAMC/1079	Item	<b>Suggestion book from the Officers' Mess of the 15th (Scottish) General Hospital in Cairo</b>	1941-1945
7	RAMC/1154/2/21	Piece	<b>Letter from No. 23 (Scottish) General Hospital, asking for information re brain injuries</b>	14 March 1942
8	SA/CSP/C.2/4/4	Item	<b>Scottish Physiotherapy Hospital and School</b>	1968-1970
9	SA/MWF/A.1/1	Item	<b>Council and General Meetings</b>	May 1916-Jul 1924
10	SA/MWF/C.9/4	Piece	<b>Ella Hepworth Dixon, 'A Modern Woman - Mrs Kinnell' (re her involvement with Royal Free, and Scottish Women's Hospital in Serbia (xerox), with letter from Edith (Gilchrist) to Joan (Haram), 14 Aug 1980</b>	1980
11	SA/QNI/H.2/3	Item	<b>Memorandum from the Institute Treasurer to the President and Council arguing against the connection with St Katherine's Hospital [Found amongst Scottish branch financial policy records, S.1/9]</b>	1902
12	SA/QNI/S.1/11	Item	<b>Statement by the Council of the Queen Victoria's Jubilee Institute for Nurses (QVJIN) Scottish Branch on the extent of District Nursing available in Scotland and suggestions for relief of demand for hospital accommodation</b>	1925

Figure 4: Results surfaced by a free-text search across the Wellcome Library catalogues

To return to the *Codebreakers: Makers of modern genetics* - users making use of digitised archives from the Wellcome Library collections navigate a highly orthodox hierarchy, beginning at a detailed fonds level description and working down to more concise file-level descriptions. Digital materials are delivered at file level – with all material in a particular file delivered as an individual “item”. A typical example from the Francis Crick papers might be the file with the reference number PP/CRI/J/1/1/18/1, whose title is simply “Correspondence: W”. It is one of 16 files in the fonds with this title (Figure 5).

This is the result of the application of the principle of original order to the fonds during arrangement – Crick’s office staff ordered his correspondence in annual files in alphabetical order of the correspondent’s surname. In isolation from a list showing the complete hierarchical arrangement of the fonds, it is impossible to know what this file relates to. Whose correspondence is it? Are there also correspondence files for A-V and X-Z? The user cannot determine which file a particular correspondent’s letter *would* be in, or if such letters exist, unless they work through the entire hierarchy. Once a user does find the desired digital image – by working through the images in the file individually, in the order determined by the arrangement – then it is impossible to find related material without clicking back into the hierarchy all over again and resuming navigation there. When acquired Crick’s papers were seen as a priority by the Wellcome Trust, and unusually correspondents’ names were included in the *Description* field. Therefore, a word-search can find them; enabling readers to sew together the two sides of a correspondence – but it is a laborious process. In a file of correspondence by Crick, the reader looking at a letter to James Watson (his co-discoverer of DNA) does not necessarily want to see next the letter Crick sent shortly afterwards to someone else, whose surname began with “W”, but the letter Watson sent to Crick in reply. The finding-aid data does not, in its current form, enable such lateral links across the hierarchy or other fonds to e.g. identify correspondence between Crick and Watson. Such lateral searching would involve time consuming meta-data enhancement to both improve titles, create indexes and include ISAD(G)’s *Name of creator* element (mandatory since the publication of ISAD(G) version 2 in 2000).

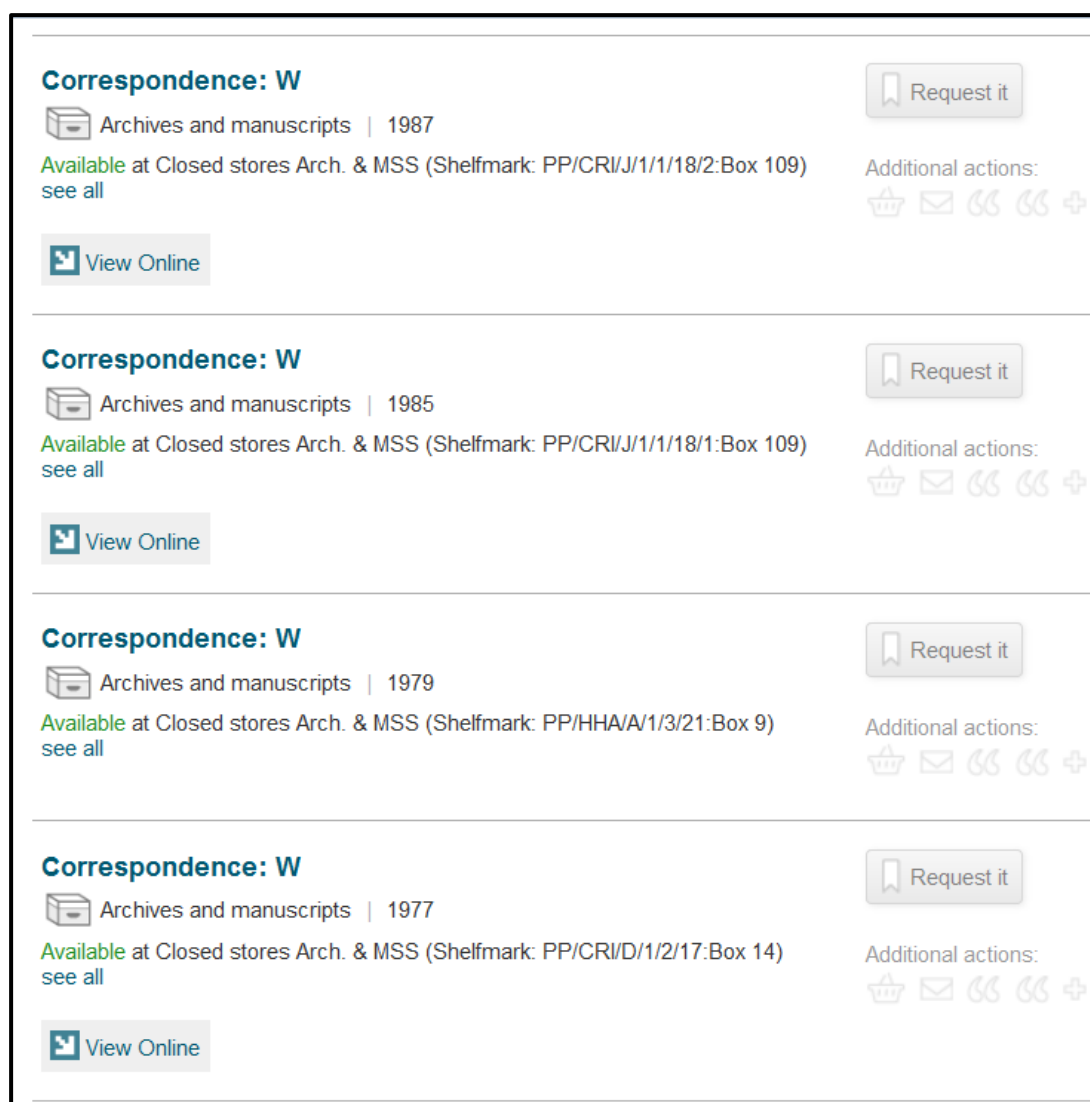


Figure 5: A selection of file level descriptions entitled “Correspondence W” in the Francis Crick Papers (PP/CRI) held at the Wellcome Library

### Access in the digital age

Crucially, users of online archives catalogues, such as those described above, now come to them primed by their experiences of other online information resources from outside the sector, accessed through the same screen – possibly even in the same sitting.

Inevitably our catalogues will be compared with those resources in terms of information-richness and usability. These issues are enmeshed: a catalogue that is not user-friendly will be perceived as low in useful information if the information, albeit present, cannot be found. It is, therefore, appropriate to look at some of those other resources that enable searching and browsing across a collection: three very widely-used applications have been chosen as examples – all notably also enable the user to interact with the data and the site: Amazon, Facebook and Flickr.

Amazon, whilst now a route to purchase almost anything, has its roots in the book trade and has at its heart a catalogue of books similar to a library catalogue. Fields holding basic bibliographic data (author, title, publisher and publication date, ISBN, physical format etc) can be cross-searched through a single search box. The search can then be refined by selecting from a series of classification lists. Once an item has been selected there are different ways to continue browsing through the database: clickable links allow users to view works by the same creator (although authors with the same name are not reliably

disambiguated) or to view works with the same classification. It is possible, for many titles, to browse through digitised images of some of the book's pages. Users can interact with the data by linking together their chosen items to form a wish-list, and, if desired, share this through e-mail, Facebook or Twitter. They can also review items (using their real name or a nickname) and browse across the reviews (and the wish-lists) created by another individual. In addition to a user's individual interaction with the metadata, recommendations regarding others purchasing choices are made based on this interaction. An individual's search history is used to identify what other viewers of a particular page went on to view or to purchase. A crude search, then, is of less relevance to the users than the possibility of an onward journey from a particular page and serendipitous finds through cross-searching, interactivity and recommendations.

Facebook's handling of images is another instructive comparison. Members have the option of both posting photographs directly to their *Timeline* of status updates and of creating albums with their own metadata. This creates an hierarchical arrangement of material, with the Facebook member's profile representing the *fonds*, the album a sub-*fonds* or equivalent, and the individual image a producible item. As with an archival hierarchy, descriptive metadata can be placed either at album or image level. Navigation of these resources, however, is achieved not through searching the metadata but entirely through linking. It is not possible to search the text of image descriptions to see if the city of Lincoln is mentioned; however, if the person posting a photograph has tagged it "Lincoln" it will show up as a marker on their map of places visited. This enables access to all their photographs tagged "Lincoln" and Lincoln's own Facebook page – and from that, in theory, to the pages of every person who has tagged themselves (or been tagged) as having been there, although many will be set to share nothing but their name with people outside their friend-list. Likewise, if a photograph is tagged as including a given person, it can be shared so that it appears on that person's own profile page and within their own collections of photographs. Material thus appears not merely in the "fonds" of its creator, but also in the "fonds" of those people who are its subjects. Metadata about a photograph's creation date enables yet another route into the collection, with images appearing on an individual's *Timeline*. Description metadata accompanies a photograph, but plays virtually no role in its discovery or in determining or facilitating an onward journey. What is important is the web of links within which a given image is located, and which provide a large range of possible onward journeys. From a given photograph browsing might include: forward to the next image, back to the previous one, up to the album level or up to the *fonds* level represented by the *Profile* of the person posting it. This corresponds to normal archival hierarchical browsing – but other links are also provided to: associated *Profiles* (and from there to whatever their profile offers: more images, their links, status updates, friends and so forth), and places tagged in the image (and from there to other photos taken in those places or the places' own Facebook pages). Thin and largely uncontrolled metadata is compensated for by linking and a wide choice of onward journeys.

Flickr offers a similar range of functionalities for image sharing. Users post images that can be arranged in a number of ways: as collections of thematic albums, as individual albums, or as a single *photo-stream* presented in date order. These correspond to the hierarchical arrangement of the *fonds* while making alternative views possible. Curated galleries can also be created – allowing individual contributors to showcase their best work. Images can also be browsed through a map interface, date lists (called *Archives*) and user tags. Searches can be restricted to an individual's contributions or made across the whole dataset. Free text searches scour user created descriptive metadata and user assigned *tags*. Tagging, unlike structured word lists, classification schemes or thesauri does not disambiguate terms, so that a search for "Lincoln" brings up photographs of Lincoln the city, Lincoln the car, or Abraham Lincoln. Advanced searching, with Boolean logic, can be used where such disambiguation is required although this functionality is not obvious to the casual user. Rights information can also be searched, enabling users to find images which can be legally used in other

implementations. Meanwhile technical metadata regarding the equipment used, the location and the date of capture are automatically created on upload. The user also has opportunities to interact – they may add images to their favourites (a self-curated album that appears on their own profile), comment on images or share them through Facebook, Pinterest, Twitter or Tumblr. Users can also join a community by joining a *Group*.

Three resources that many archive users will be familiar with before their interactions with the archive catalogue begin. Each has its own characteristics, and each has a different purpose – there is a contrast, for instance, between Amazon's commercial drive and the other two's social, sharing function. There are, however, key features that each has in common, in which they contrast with archive and library catalogues. Firstly, the descriptive metadata is typically lightweight, and either uncontrolled, or at the very least unreliable (e.g. failures in disambiguation in author names in Amazon or user assigned tags in Flickr). The data may be searchable, but the main search mechanism is a simple key-word search. There is arrangement in a fashion that archivists and information professionals would understand – books can be grouped by author, or images by creator – but this is not the only system of arrangement: it is possible to group items in all sorts of other ways, and in ways that are emergent, not pre-ordained, and unpredictable, arising out of users' own interactions with the material. A key point is the multiplicity of onward journeys: each item included in these resources offers a wide variety of onward links, a variety of connections and arrangements rather than tying the user to one authorised arrangement, one authorised set of connections.

None of these resources offer what information professionals would consider high-quality, controlled metadata. Their popularity presumably means that users value other properties in information resources: the interconnectedness of the material, the multitude of ways items can be linked together, the opportunity to shape those linkages for their own purposes and the chance to interact with others about the materials.

### **Rethinking archival arrangement and description**

In contrast to the popular resources described above, archival finding-aids, which comply with archival principles, have been dominated by a belief in one correct arrangement to the exclusion of others. The relationship between the needs of physical storage and the intellectual arrangement is implicit in archival principles – an intellectual file can be packaged as a physical file and produced for the user to consult. In the digital realm, where the storage location is irrelevant, centring on the fonds as the organising principle, above all others obscures other possible arrangements: relationships, interactions, timelines and themes. These dynamic qualities of archives are precisely what users are likely to be seeking. Meanwhile the descriptions created often divorce the metadata from the data itself, by describing aggregations of material rather than individual items. Hierarchical catalogues provide only a rough signposting of where something *might* be, but are surfaced by powerful search engines that take the user directly to the word searched for, often divorced from the context of the hierarchy. Generalised aggregate description can leave little trace of the details of the content so that such finding-aids are essentially unfindable. In a world where the use of search engines is habitual and unthinking, if description leaves no trace of a particular person or activity, much material is effectively hidden to most users (Gwinn-Becker, 2014) – as the recent case of the “discovery” of a medical report concerning the death of Abraham Lincoln brought into sharp focus (Fischer, 2012; Papaioannou, 2012).

The archives profession now faces a clear dilemma in the application of archival principles and the cataloguing standards that conform to these. Processing backlogs have seen the archives profession embracing *more product, less process* as a way of reducing these. The adoption of this methodology complies with archival principles, but sees arrangement and description taking place at the aggregate level – principally the series or file. Meanwhile, for digitised material the contextual hierarchical arrangement is less important than the need for

it to be adequately discovered. For this it needs item level description with indexing or tagging, and links to related materials to enable cross-linking and lateral browsing – more akin to the labour intensive document calendaring seen in the early days of the profession.

Additionally, to be Web 3.0 ready archival descriptions need to be *linked data ready* through the use of relevant ontologies and vocabularies so that “archives can benefit from being presented not in isolation, but as part of all the diverse data sources that can be found” (Linking Lives Project, n.d.). Item level description, and in particular indexing has a considerable resource implication. But the profession does need to consider what constitutes a “good-enough” standard and see if it can deliver what the user needs.

Good cataloguing makes archival material findable and digitised facsimiles enable remote users to replicate the experience of the reading room. But the extra possibilities of the digital environment – search engine discovery, linkages across collections to related material, user arrangement and a merging of the description and digital resource - provide the user with an onward journey. Hierarchical arrangement and sketchy catalogues is a tired model that sees all the advantages of online delivery, seized triumphantly by the new web-based resources, left to one side. There is a risk, if we do not examine the new ways in which readers consume resources in this environment, that our “correct” hierarchical catalogues will be nothing more than carefully curated silos of irrelevance, laying out data in accordance with a set of rules that may well reflect our ultimate principles, but which fails to deliver a useful reader experience.

After nearly twenty years of the real-world piloting of ISAD(G), it is now appropriate to assess, as a profession, the extent to which it has met the needs of our users and the areas in which further work may be required.

### **User study**

The authors propose a study that aims to identify what arrangement and description users need to enable them to discover and use digital facsimiles of archival material in an online environment, and whether these processes can be effectively crowd-sourced. Taking a user perspective, the objectives of the study will be to: research the status of archival description across the profession; establish the usability and functionality of existing online catalogues, in comparison with popular online resources; to identify what would constitute a “good enough” archival cataloguing standard to satisfy the user, professional and current and future technical perspectives; and to explore the feasibility of crowdsourcing description and user generated arrangement.

Using a UK lens, the study will firstly establish the dimensions and processes of archival description. It will examine the cataloguing policies and processes of archives repositories including: the uptake of archival standards, how they have been implemented, how much resource is available for cataloguing tasks, what percentage of catalogues have been retro-converted, the online presence of archives (both catalogues and digital facsimiles), the processing backlog and how collections are selected for processing from this.

Focussing on catalogues which also offer digital facsimiles, the above investigations will inform the development of a study to establish the user perspective of existing online catalogues: how they are navigated; how successful searches are; what could improve searching and browsing, and required and desirable functionality. Parallel investigation will be undertaken regarding the technical requirements for ensuring that catalogues can be adequately searched and browsed in both the current and developing Internet paradigms.

These studies will help to identify what is needed to develop a “good-enough” archival descriptive standard, along with the development of a test environment in which related



digitised archival collections will be Web mounted, with minimal contextual metadata and no pre-defined arrangement. Users will be invited to describe the collections, make appropriate links to related resources and arrange them as they see fit, using the provided tools. User behaviour will be analysed to evaluate whether crowdsourcing can be used to realise the needs of competing realities for the profession: the imperative of the pared down working recommendations of *more product, less process* and the need for detailed item level cataloguing to enable digital facsimiles to be discovered.

## Conclusion

The development of ISAD(G) and related archival standards, rooted in archival principles, represented a *great leap forward* for the archives profession. Appropriate methodologies and software, along with their high level of take-up, enabled archival descriptions to be created using a consistent hierarchical structure for: mounting on the Internet, sharing in networks, and attaching to digital facsimiles. However, little has changed in archival descriptive practice since the mid-2000s while use and access to the Internet has changed beyond recognition, and continues to evolve. A reassessment of the practice and processes of archival descriptive standards, with an emphasis on the user perspective, will help to establish how archival collections can become more readily discovered both now and in the future.

## Bibliography

Ad Hoc Commission on Descriptive Standards (1994). *ISAD(G): General International Standard Archival Description*. Stockholm.

Adlib Information Systems (1999). "For flexible archive management".  
<<https://web.archive.org/web/19990208021107/http://www.uk.adlibsoft.com/>>  
[Consulted: 03/09/2014].

AIM25 (2013). "AIM25: Archives in London and the M25 area". <<http://www.aim25.ac.uk/>>  
[Consulted 08/09/2014].

Archives Hub (n.d.-a). "Archives Hub". <<http://www.archiveshub.ac.uk/>> [Consulted: 08/09/2014].

Archives Hub (n.d.-b). "What is Genesis?" In: *Archives Hub*.  
<<http://archiveshub.ac.uk/features/genesis/>> [Consulted 08/09/2014].

Archives Task Force (2004). *Listening to the past, Speaking to the future: Report of the Archives Task Force*. London: Museums, Libraries and Archives Council.

Barrit, Marjorie Rabe (1993). "Coming to America: Dutch archivistic and American archival practice". *Archival Issues*. n. 18, v. 1, p. 43–53.

Bouché, Nicole (1997). "Implementing EAD in the Yale University Library". *American Archivist*. n. 60, p. 408–419.

Bunn, Jennifer (2013). "Developing descriptive standards: a renewed call to action". *Archives and Records*. n. 34, v. 2, p. 235–247.

Craven, Louise (2004). "Linking arms". <<http://www.ukat.org.uk/news/lcraven20040616.pdf>>  
[Consulted 08/09/2014].



- DS (1999). "CALM 2000 for Archives upgrade".  
<<https://web.archive.org/web/19990128121417/http://www.dsltd.co.uk/products.htm#CALMarc>> [Consulted 03/09/2014]
- Duff, Wendy (1995). "Will metadata replace archival description : A commentary". *Archivaria*. n. 39, p. 33–38.
- Fischer, Suzanne (2012, June 19). "Nota bene: If you "discover" something in an archive, it's not a discovery". In: *The Atlantic*.  
<<http://www.theatlantic.com/technology/archive/2012/06/nota-bene-if-you-discover-something-in-an-archive-its-not-a-discovery/258538/>> [Consulted 07/09/2014].
- Genesis Project Team (2004). "Genesis: Developing access to women's history sources in the British Isles".  
<<http://www.webarchive.org.uk/wayback/archive/20060404120000/http://www.genesis.ac.uk/index.html>> [Consulted 08/09/2014].
- Greene, Mark, & Meissner, Dennis (2005). "More product, less process: Pragmatically revamping traditional processing approaches to deal with late 20th-century collections". *The American Archivist*. n. 68, p. 208–263.
- Gwinn-Becker, Kirsten (2014). "Kill the finding aid! Give the public the ability to really search our history". In: *Giving history a future Let's talk about innovative solutions and technologies to deliver robust digital archives*. <<http://blog.historyit.com/kill-finding-aid/>> [Consulted 07/09/2014]
- Higgins, Richard (1998). "A case study of EAD implementation at Durham University Library Archives and Special Collections". *Archives and Museum Informatics*. n. 12, p. 221–234.
- Higgins, Sarah, & Inglis, Gavin (2003). "Implementing EAD: the experience of the NAHSTE project". *Journal of the Society of Archivists*. n. 24, v. 2, p. 199–214.
- Hill, Amanda, Stocking, William, & Higgins, Sarah. (2004). "Different strokes for different folks: Presenting EAD in three UK online catalogues". In *Encoding Across Frontiers: Proceedings of the European Conference on Encoded Archival Description and Context (EAD and EAC), Paris, France 7-8 October 2004* p. 183–195.
- International Council on Archives (1992a). "ISAD(G): General international standard archival description". *Archivaria*, n. 34, p. 17–32
- International Council on Archives. (1992b). "Statement of principles regarding archival description". *Archivaria*. n. (34), p. 8–16.
- International Council on Archives (2000). "ISAD(G): General international standard archival description". In: *International Council on Archives Committee on Descriptive Standards* Retrieved from <[http://www.icacds.org.uk/eng/ISAD\(G\).pdf](http://www.icacds.org.uk/eng/ISAD(G).pdf)> [Consulted 08/09/2014]
- International Organization for Standardization (2001). *ISO 15489-1:2001, Information and Documentation - Records Management - Part 1: General*.
- Jenkinson, Hilary (1966). *A manual of archive administration* (2nd ed.). London: Lund Humphries.

- Johnston, Dorothy (2001). "From typescript finding aids to EAD (Encoded Archival Description). A university case study". *Journal of the Society of Archivists*. n. 22, v. 1, p. 39–52.
- Library of Congress. (1999). "EAD application guidelines for version 1.0" In: *Encoded Archival Description version 2002 official site*.  
<<http://www.loc.gov/ead/ag/agcontext.html>>. [Consulted 07/09/2014]
- Linking Lives Project (n.d.). "About linking lives".  
<[http://archiveshub.ac.uk/linkinglives/?page\\_id=2](http://archiveshub.ac.uk/linkinglives/?page_id=2)>. [Consulted 07/09/2014]
- MacNeil, Heather (1994). "Archival theory and practice: Between two paradigms". *Archivaria*. n. 37, p. 6–18.
- MacNeil, Heather (1995). "Metadata strategies and archival description: Comparing apples to oranges". *Archivaria*. n. 39, p. 22–32.
- MacNeil, Heather (2002). "The context is all: Describing a fonds and its parts in accordance with the Rules for Archival Description". In: Eastwood, Terry (ed.). *The archival fonds: From theory to practice*. Ottawa: Bureau of Canadian Archivists, p. 195-229.
- McCrea, Donna (2006). "Getting more for less: Testing a new processing model at the University of Montana". *The American Archivist*, 69(Fall-Winter), 284–290.
- Meehan, Jennifer (2009). "Making the leap from parts to whole: Evidence and inference in archival arrangement and description". *The American Archivist*. n. 72, v. 1, p. 72–90.
- Muller, Samuel, Feith, Johan, & Fruin, Robert (1968). *Manual for the arrangement and description of archives drawn up by the direction of the Netherlands Association of Archivists* (2nd ed.) New York: Wilson.
- National Library of Wales (2014). "Archives Wales: Your gateway to history".  
<<http://www.archiveswales.org.uk/>> [Consulted 08/09/2014]
- Papaioannou, Helen (2012, June 21). "Actually, yes, it \*is\* a discovery if you find something in an archive that no one knew was there". In: *The Atlantic*.  
<<http://www.theatlantic.com/technology/archive/2012/06/actually-yes-it-is-a-discovery-if-you-find-something-in-an-archive-that-no-one-knew-was-there/258812/>> [Consulted 08/09/2014]
- Planning Committee on Descriptive Standards (2008). *Rules for archival description*. Ottawa, Canada.
- Roberts, Alwyn (2014). *Creating access to archives*. Aberystwyth.
- School of Oriental and African Studies (n.d.). "Mundus: Gateway to missionary collections in the UK". <<http://www.mundus.ac.uk/about.html>> [Consulted 08/09/2014]
- Scottish Archive Network Ltd. (n.d.). "Welcome to the Scottish Archive Network".  
<<http://www.scan.org.uk/>> [Consulted 08/09/2014]
- Society of American Archivists. (2013). *Describing archives: A content standard* (2nd ed.). Chicago: Society of American Archivists.

Society of American Archivists, & Library of Congress (1998). "EAD tag library for version 1.0". In: *Encoded archival description version 2002 official site*.  
<<http://www.loc.gov/ead/tglib1998/index.html>> [Consulted 08/09/2014]

Sweet, Meg (2001). "The internationalisation of EAD (Encoded Archival Description)".  
*Journal of the Society of Archivists*. n. 22(1), p. 33–38.

Wallace, David (1995). "Managing the present: Metadata as archival description". *Archivaria*.  
n. 39, p. 11–21.

Wellcome Library (2014). "Codebreakers: Makers of modern genetics".  
<<http://wellcomelibrary.org/using-the-library/subject-guides/genetics/makers-of-modern-genetics>> [Consulted 08/09/2014]

Yeo, Geoffrey (2012). "Bringing things together: Aggregate records in a digital age".  
*Archivaria*. n. 74 (Fall), p. 43–91.